



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁷ : H03M 13/00</p>	A2	<p>(11) International Publication Number: WO 00/57560</p> <p>(43) International Publication Date: 28 September 2000 (28.09.00)</p>																												
<p>(21) International Application Number: PCT/GB00/01068</p> <p>(22) International Filing Date: 22 March 2000 (22.03.00)</p> <p>(30) Priority Data: 9906588.0 22 March 1999 (22.03.99) GB</p> <p>(71) Applicant (for all designated States except US): LUCENT TECHNOLOGIES INC. [US/US]; 600 Mountain Avenue, Murray Hill, NJ 07974-0636 (US).</p> <p>(72) Inventors; and (75) Inventors/Applicants (for US only): BENEDETTO, Sergio [IT/IT]; Via Assarotti, 7, I-10122 Torino (IT). DRAKUL, Spase [SI/GB]; 75 Berkeley Road, Wroughton, Swindon, Wiltshire SN4 9BN (GB). MASERA, Guido [IT/IT]; Via Martini, 5, I-10025 Pino Torinese (IT). MONTORSI, Guido [IT/IT]; Strada Valsalice, 19, I-10131 Torino (IT). PICCININI, Gianluca [IT/IT]; Strada Valle Fet, 15, I-10090 Gassino Torinese (IT). ZAMBONI, Maurizio [IT/IT]; Corso Duca Degli Abruzzi, 98, I-10129 Torino (IT).</p> <p>(74) Agents: WILLIAMS, David, J. et al.; Lucent Technologies UK Limited, 5 Momington Road, Woodford Green, Essex IG8 OTU (GB).</p>																														
<p>(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p style="text-align: center;">Published</p> <p style="text-align: center;"><i>Without international search report and to be republished upon receipt of that report.</i></p>																														
<p>(54) Title: IMPROVED TURBO DECODER 1</p> <p>(57) Abstract</p> <p>In a radio telecommunications system operating an iterative version of the MAP algorithm, a decoder for a data source or data receiver comprising three memories (80, 82, 84) arranged to perform in parallel the steps of updating the NDP alpha metrics in a forward direction; updating the NDP beta metrics in a backward direction, and computing NDP values of the output probability distribution. The arrangement can be regarded as a split RAM SISO architecture.</p>																														
<div style="display: flex; align-items: center;"> <div style="margin-right: 20px;"> <p style="text-align: center;">80 ↓ RAM 1</p> <p style="text-align: center;">82 ↓ RAM 2</p> <p style="text-align: center;">84 ↓ RAM 3</p> </div> <table border="1" style="border-collapse: collapse; text-align: center;"> <thead> <tr> <th></th> <th style="width: 33%;">RAM 1</th> <th style="width: 33%;">RAM 2</th> <th style="width: 33%;">RAM 3</th> </tr> </thead> <tbody> <tr> <td style="text-align: right;">phase 1</td> <td>S1</td> <td></td> <td></td> </tr> <tr> <td style="text-align: right;">phase 2</td> <td>S2 {B1 (init)}</td> <td>S1</td> <td></td> </tr> <tr> <td style="text-align: right;">phase 3</td> <td>S3 B2 (init)</td> <td>S2 A1</td> <td>S1</td> </tr> <tr> <td style="text-align: right;">phase 4</td> <td>S4 B3 (init)</td> <td>S3 A2</td> <td>S2 B1 P1</td> </tr> <tr> <td style="text-align: right;">phase 5</td> <td>S5 B4 (init)</td> <td>S4 A3</td> <td>S3 B2 P2</td> </tr> <tr> <td style="text-align: right;">phase 6</td> <td>S6 B5 (init)</td> <td>S5 A4</td> <td>S4 B3 P3</td> </tr> </tbody> </table> </div>				RAM 1	RAM 2	RAM 3	phase 1	S1			phase 2	S2 {B1 (init)}	S1		phase 3	S3 B2 (init)	S2 A1	S1	phase 4	S4 B3 (init)	S3 A2	S2 B1 P1	phase 5	S5 B4 (init)	S4 A3	S3 B2 P2	phase 6	S6 B5 (init)	S5 A4	S4 B3 P3
	RAM 1	RAM 2	RAM 3																											
phase 1	S1																													
phase 2	S2 {B1 (init)}	S1																												
phase 3	S3 B2 (init)	S2 A1	S1																											
phase 4	S4 B3 (init)	S3 A2	S2 B1 P1																											
phase 5	S5 B4 (init)	S4 A3	S3 B2 P2																											
phase 6	S6 B5 (init)	S5 A4	S4 B3 P3																											

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

IMPROVED TURBO DECODER 1

5

This invention relates to a decoder, that is a decoder used in a radio telecommunications system in which digital data has been coded by turbo code to improve performance, and relates especially to a de-interleaver in the decoder.

One type of code which is applied in radio telecommunications systems is a concatenated convolutional code ("turbo code") which implements the iterative version of the well known Maximum A Posteriori (MAP) algorithm. The core of the MAP decoding algorithm is a procedure to derive the sequence of probability distributions over the information symbols alphabet based on the received signal and constrained on the code structure. The MAP algorithm is described by L R Bahl, J Cocke, F Jelinek and J Raviv: "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate". IEEE Transactions on Information Theory, pp. 284-287, March 1974. Efficient implementation of the algorithm in an integrated circuit is not easy, and in some versions the algorithm is modified.

In one version, logarithmic approximation is used, which renders the algorithm additive by replacing all metrics with logarithmic qualities, as described by S Benedetto, D Divsalar, G Montorsi and F Pollara, "Soft-output decoding algorithms for continuous decoding of parallel concatenated convolutional codes", Proceedings of ICC'96, Dallas, Texas, June 1996 and S Benedetto, D Divsalar, G Montorsi and F Pollara, "Soft-input soft-output modules for the construction and distributed iterative decoding of code networks", European Transaction on Telecommunications, vol. ETT 9, March/April 1998.

In another version, also described in the two references just given, a sliding window version of the algorithm is provided, where the decoder operates on a fixed memory length, instead of requiring that the whole transmitted sequence is stored.

Even in these simplified versions, the decoding algorithm remains critical for high rate applications. The requirement of performing a substantial number of decoding iterations while supporting a high data rate can pose serious implementation problems, especially for an implementation in low cost ASICs.

Conventionally, a complete turbo decoder usually includes two kinds of blocks; the first kind comprises Soft Input Soft Output (SISO) stages which implement the

MAP algorithm, and the second kind comprises de-interleavers, which scramble (de-scramble?) the processed data according to the interleaving laws used in the encoder. Other kinds of blocks may also be required for the implementation of the decoder, such as RAM memories for storing data through the iterations or synchronization circuits.

- 5 These blocks can be inter-connected in the decoder in many topologies, as described in the paper in ETT Volume 9 given above. Two alternative decoding strategies can be adopted: if a single instance for each SISO stage is allocated, the required iterations are performed serially (serial decoding) and the overall decoding speed is N_{it} times slower than the speed of the SISO where N_{it} is the number of iterations. As an alternative, a
10 parallel decoding can be adopted, where multiple SISO stages are allocated for each iteration and the global decoding speed is equal to that of the SISO stage. In this case, a feed-forward is used, consisting of a chain of blocks that process the data in pipelining.

An important role is played in the decoder by the interleaver and the de-interleaver, which usually add a large contribution to the whole cost, mainly for three
15 reasons:

- 1 the error rate performance of the code is inversely proportional to the interleaving length, thus the interleaver implementation involves a large RAM;
- 2 the size of the memory devoted to the storage of the reliabilities
20 computed by the de-mapper is also related to N_{int} , the Number of Interleavings;
- 3 it has been proved that, at least for high signal to noise ratios, the best performances are obtained with randomly generated interleaving patterns: this means that the required sequence of addresses cannot be obtained through simple computations but quite large ROMs must be allocated.

25 It is an object of the invention to reduce the complexity of a decoder circuit.

According to the invention, a method of decoding concatenated convolutional codes comprising the steps of storing a block of $N \cdot DP_{input}$ branch probabilities in a memory; from the stored block:

- updating NDP alpha metrics in a forward direction;
- 30 updating NDP metrics in a backward direction; and
- computing NDP values of the output probability distributions;

characterised in that the two updating steps and the computing step are performed in parallel. NDP is the length or width of the sliding window.

Also according to the invention, a radio telecommunications system in which each data source and data receiver comprises a digital coder/decoder for applying concatenated convolutional codes, characterised by a decoder comprising three
5 memories each having a single read-modify-write access, the memories being arranged to perform in parallel the steps of updating NDP alpha metrics in a forward direction; updating NDP beta metrics in a backward direction; and computing NDP values of the output probability distributions.

10 The code may be an iterative version of the Maximum A Posteriori Algorithm.

The prior art will be described with reference to Figures 1, 2, 3 and 4a of the accompanying drawings in which:-

Figure 1 illustrates schematically a part of a radio telecommunications system;

Figure 2 illustrates schematically a part of the encoding/decoding and
15 modulating/demodulating part of the radio telecommunications system;

Figure 3a illustrates schematically a more detailed view of a convolutional encoder and 3b is a more detailed view of a convolutional decoder.

The invention will be described with reference to Figures 4, 5 and 6 in which :-

Figure 4 illustrates the timing of memory operations by the use of three RAMs;

20 Figure 5 illustrates the decoder architecture incorporating three RAMs; and

Figure 6 illustrates a modified version of Figure 5.

In Figure 1, a radio telecommunication system comprises a Core Network (CN)
10 having an interface 12 with a Radio Access Network (RAN) 14 which in turn has an interface 16 with a plurality of mobile users 18, 20. In the RAN 14 are two Base
25 Station Controllers (BSC) 22, 24 each controlling two Base Transceiver Stations (BTS) 26, 28, 30, 32. In practice there will be many BSCs controlling many BTSs.

Referring now to Figure 2, in a transmitter 40 a data source 42 supplies data to an encoder 44 which applies the MAP algorithm to the data; the encoder output is connected to a modulator 46 which modulates the encoder data signal, and supplies it
30 to a communications channel 48. In a receiver 50, a demodulator 52 receives a signal from the channel 48, demodulates it and passes the signal to a decoder 52. The

decoder 52 decodes the data signal and supplies it to user equipment 54, such as one of the mobiles 18, 20.

While traversing the channel 48 the signal is subject to noise N, to fading F, and to frequency offset O; to ameliorate these deleterious effects, coding such as turbo
5 coding is used.

Figure 3a illustrates an encoder 44. An incoming stream of bits N passes to a first buffer 60 where the stream is divided and passes to a first convolutional encoder 62 and also through an interleaver 64 and a second buffer 66 to a second convolutional encoder 68. The first encoder operates at rate $\frac{1}{2}$ and provides two encoded output
10 signals X1 and X2. The second encoder operates at rate 1 and provides a third output signal X3.

Figure 3b illustrates a decoder 70, comprising a first decoder core element 72 operating on SISO principles, an interleaver 74, a second decoder core element 76 also operating on SISO principles, and a de-interleaver 78. The signal X1 is supplied to the
15 first decoder element 72 and the signal X2 is supplied to the second decoder element 76. The third signal X3 is supplied to both decoders. The output of the de-interleaver is a decoded data signal.

Each mobile 18, 20 in Figure 1 and each BTS 26, 28, 30, 32 is provided with an encoder and a modulator, and also with a decoder and a de-modulator.

20 Reference has been made above to the use of the sliding window version of the MAP algorithm, in which the decoder operates on a fixed memory length. An improvement proposed by A J Viterbi, "An intuitive justification of the MAP decoder for convolutional codes", IEEE Journal on Selected Areas in Communications, vol. 16, no. 2, Feb. 1998 as an alternative to the storage of the entire stage metric history.

25 The basic idea of this solution is to double the extension of the sliding window from N... D... P... (NDP) to 2 NDP: in this way, after the initial NDP steps of the backward recursion required by the algorithm, the computed beta values have a memory longer than NDP (in the sense that they include the contribution of more than NDP branch metrics through the trellis) and can therefore be used to directly to feed an
30 output.

To describe the proposed SISO architecture, let us define as S the operation of

storing a block of NDP input branch probabilities in a RAM; according to the MAP algorithm, this block of written values must be read three times for performing the following operations:

- 1 Operation A, which is the updating of NDP alpha metrics in the forward
5 direction;
- 2 Operation B, which is the updating of NDP beta metrics in the backward
direction;
- 3 Operation P, which is the computing of NDP values of the output
probability distribution.

10 In the inventive arrangement, these three operations are performed in parallel. In order to avoid the use of a multiple port RAM, three separated memories of depth NDP are used so that the defined operations operate on them in a cyclic way. One additional memory is required for temporarily storing the computed values of alpha.

Figure 4 illustrates the inventive arrangement of three RAMs 80, 82, 84 in a
15 decoder such as 72 or 76. A sequence of 6 phases is indicated where a phase indicates a processing of a whole block of NDP input metrics.

In the figure it can be seen that the B operation is performed twice for each block: the first time beta values are updated for the first NDP steps, but are not used for the SISO output calculation; the second time the beta updating is continued for the
20 NDP following steps and NDP SISO outputs are evaluated at the same time.

1 In phase 1, the first NDP long block of branch metrics is stored in RAM
80 (S1).

2 In phase 2, the beta value is calculated on the first block (B1INIT), the
first block is moved to the second RAM 80, and a second block of branch metrics is
25 stored in the first RAM 80.

It will be apparent that read-modify-write access is required for the RAMs 80,
82, 84.

3 In phase 3 the alpha values are calculated on the first block (A1) and are
stored in a separate memory for subsequent use; the first block is moved to the third
30 RAM 84; the beta values are calculated on the second block (B2INIT) and the second
block is moved to the second RAM 82; a third block of branch metrics is stored in the

first RAM 80 (S3).

4 In phase 4 the RAM 84 content is read in the reverse order for sequentially calculating the current values of beta (B1) and the associated output probabilities (P1); this calculation also makes use of the alpha values stored in the
5 separate RAM, which can be re-used for the writing of a new alpha evaluated in the second block (A2) at the same time, the initialization is performed on the third block (B3 INIT); the second block S2 is discarded; the third block S3 is moved to the third memory 84; and a new block S4 is read into the first RAM 80.

5 In phase 5 the values B2 and P2 are calculated for the second block in
10 the third RAM 84, a value of A3 for the third block is calculated in the second memory 82, and a value B4 INIT is calculated in the first memory 80; the block S3 is discarded; block S4 is moved to the third memory 84 and block S5 is moved to a second memory 82 while a third block is introduced.

6 In phase 6 values of B3 and P3 are calculated in the third RAM 84, so
15 that all required values are available for the first three blocks. As before a new block of data is stored in the first memory 80, and other blocks are shifted along one memory then discarded.

Figure 5 illustrates the architecture of a decoder with three RAMs 80, 82, 84 each having a single input and read-modify-write access 81, 83, 85. Each RAM is
20 connected to an associated Add Compare Select (ACS) unit 86, 88, 90 operating respectively on the stages Bint; A and B as defined above, with respective input circuits 92, 94, 96. Input 92 receives three signals; input 94 receives two signals and input 96 receives two signals and the output from the ACS 86.

The output from ACS 88 associated with the second memory 82 is supplied to
25 an additional storage stage 98 which stores temporarily the computed alpha values. There is an additional ACS circuit 100, which receives signals from storage 98, from ACS 90, and from RAM 84, and provides an output signal.

(In comparison with the prior art, the ACS sections are allocated instead of an NDP long pipelining of ACS stages. This simplification does not affect the
30 performance of the decoder, as the throughput is limited by the single ACS.

The modification to the algorithm and the architecture according to the

invention provide a substantial cost reduction.

It will be appreciated that the SISO outputs are obtained in reverse order so that, to obtain the correct order in the output signal, either the interleaving law of the decoder (74 or 78 in Figure 3b) must be changed, or the decoding stage can be
5 arranged to operate on a Last In First Out principle.

Inspection of Figure 5 will show that the input probability $PI(u;I)$ and $PI(c;I)$ are added at each stage at each ACS section 86, 88, 90.

A variation of an architecture to implement the invention is shown in Figure 6. There is an additional ACS stage 102, which precedes the ACS stages 86, 88, 90 and is
10 arranged to add the input probabilities. The sum is supplied to the first memory 80. By use of such an arrangement the number of adders can be halved.

An associated requirement is that the width of each of the memories 80, 82, 84 must be increased, so that the reduction in silicon area achieved by omission of half the address is cancelled at least partly. The actual result in terms of surface area of silicon
15 must be evaluated for each individual case.

However, from the point of view of performance, the arrangement of Figure 6 always results in an increased throughput of the SISO decoder: in fact, if the adding of input probabilities is performed inside the ACS processor it increases the global ACS delay; as the ACS has the outputs connected to the input through a feedback loop, it is
20 the decoder bottleneck, and moving operations from inside to outside the ACS processor immediately gives better performance.

The embodiment has been described with reference to the Global System for Mobile Communications (GSM), but the principles described above apply equally to the Universal Mobile Telephone System (UMTS) and to the Enhanced Data Rates for
25 GSM Evolution (EDGE) system, and to any other mobile telecommunication system using any coding system, but especially turbo coding.

CLAIMS

- 1 A method of decoding concatenated convolutional codes comprising the steps of
- storing a block of NDP input branch probabilities in a memory;
- 5 from the stored block:-
- updating NDP alpha metrics in a forward direction;
- updating NDP beta metrics in a backward direction; and
- computing NDP values of the output probability distributions;
- characterized in that the two updating steps and the computing step are
- 10 performed in parallel.
- 2 A method according to Claim 1 in which for each block the updating of the NDP beta metrics is performed in two stages:-
- (a) in a first step the beta values are updated for the first NDP steps;
- (b) in a second step the beta values are updated for the
- 15 following/subsequent NDP steps and NDP SISO outputs are simultaneously evaluated.
- 3 A method according to Claim 1 or Claim 2 in which the computing steps are performed in parallel in six phases comprising:-
- in phase 1, storing (S1) a first NDP long block of branch metrics in a first memory;
- 20 in phase 2, performing an initialization (B1INIT) on said first block in the first memory, moving the first block to a second memory, and storing (S2) a second NDP long block of branch metrics in the first memory;
- in phase 3, calculating alpha values for the first block (A1) in the second memory, storing the alpha values, and moving the first block to a third memory;
- 25 performing an initialization (B2INIT) on said second block in the first memory, moving the second block to the second memory, and storing (S3) a third NDP long block of branch metrics in the first memory;
- in phase 4, calculating current beta values (B1) and associated output probabilities (P1) for the first block in the third memory, and discarding the first block
- 30 (?), calculating alpha values for the second block (A2) in the second memory and storing the alpha values, and moving the second block to the third memory; performing

an initialization (B3INIT) on said third block in the first memory, and moving the third block to the second memory;

in phase 5, calculating current beta values (B2) and associated output probabilities (P2) for the second block in the third memory and discarding the second block; calculating alpha values for the third block (A3) in the second memory and storing the alpha values, and moving the third block to the third memory; and

in phase 6, calculating current beta values B3 and associated output probability P3 on the third block in the third memory.

4 A method according to Claim 1, 2 or 3 comprising a further step of
10 deriving a decoded signal from at least the first three blocks of code.

5 A method according to Claim 2 or Claim 3 or Claim 4 in which for calculation of the current beta values B1 B2 and B3 and associated output probabilities P1 P2 P3 the contents of the third memory are read in reverse order.

6 A method according to Claim 5 further comprising the initial step of
15 storing the input probabilities $PI(u;I)$ and $PI(c;I)$ and supplying the sum to the first memory.

7 A radio telecommunications system in which each data source and data receiver (18, 20, 26, 28, 30, 32) comprises a digital coder/decoder (44, 54) for applying concatenated convolutional codes characterised by a decoder comprising three
20 memories (80, 82, 84) each having a single read-modify-write access 81, 83, 85, the memories being arranged to perform in parallel the steps of updating NDP alpha metrics in a forward direction; updating NDP beta metrics in a backward direction, and computing NDP values of the output probability distributions.

8 A decoder according to Claim 7 in which for each block of NDP input
25 branch probabilities the second memory (82) is arranged to calculate the updated alpha values from a previous NDP step in which the first memory (80) is arranged to calculate the current beta values from a previous NDP step; and the third memory (84) is arranged to calculate updated beta values for subsequent NDP steps and to calculate the output probability distributions.

30 9 A decoder according to Claim 7 in which an additional memory (98) is connected to the second memory (82) and is arranged to store temporarily values of

alpha calculated in the second memory.

10 A decoder according to any one of Claims 7, 8 or 9 further comprising an Add Compare Select means (102) arranged to add the input probabilities and to provide the sum to the first memory (80).

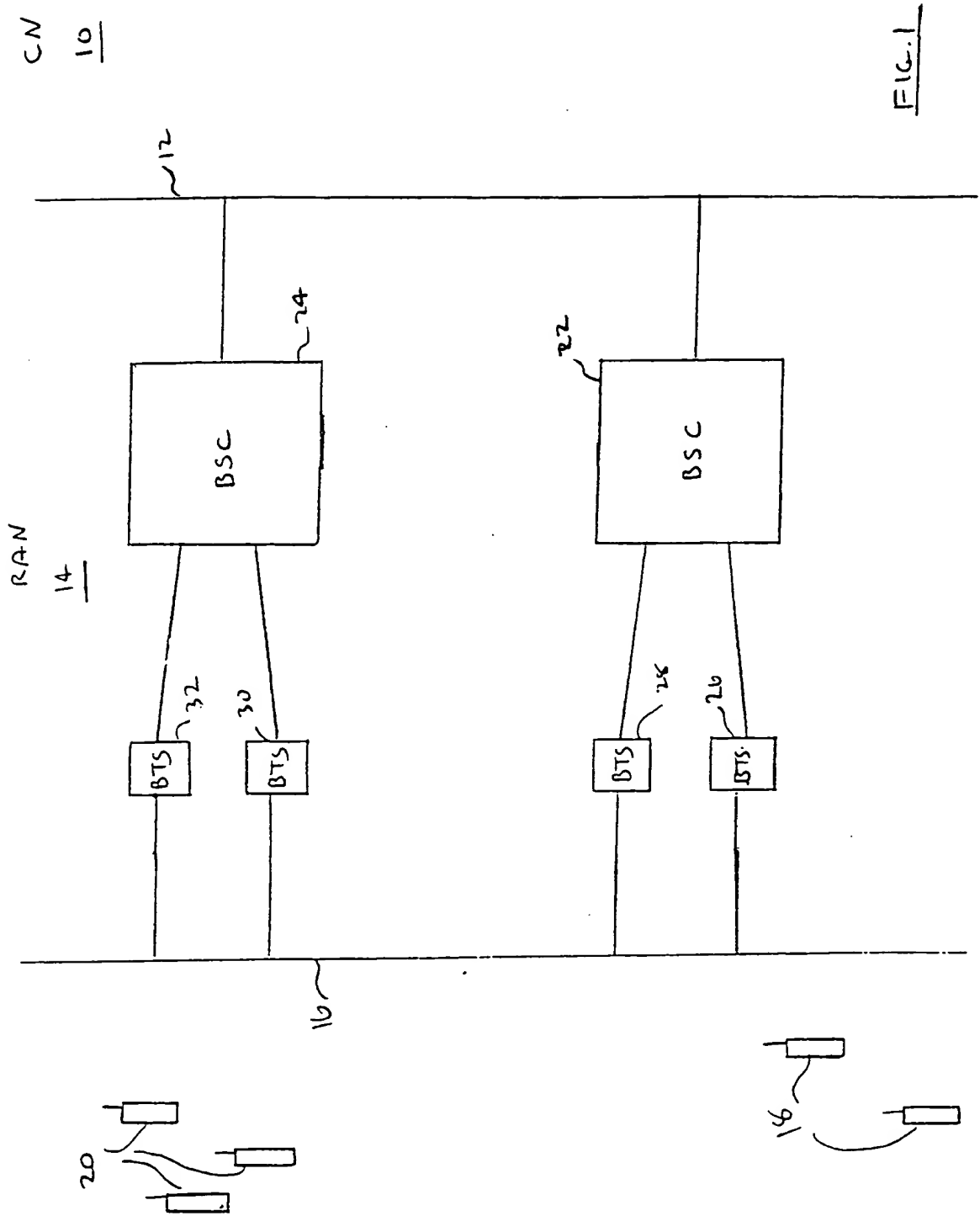
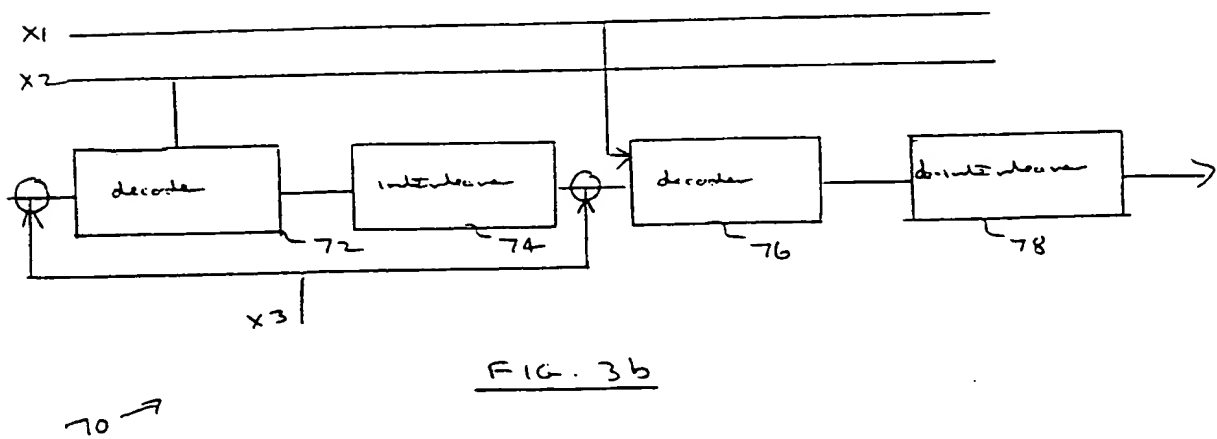
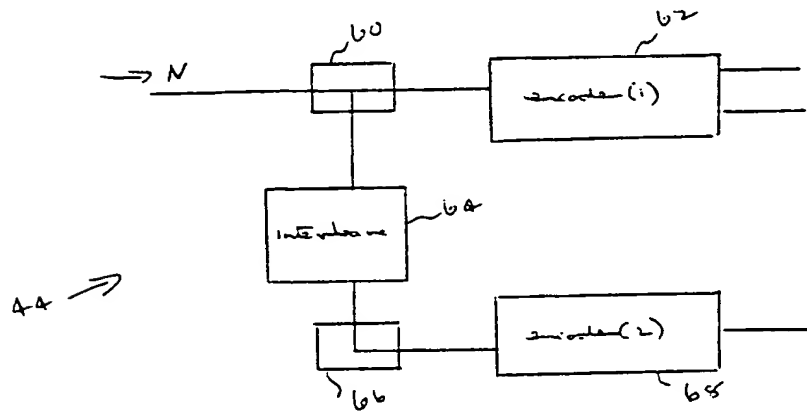
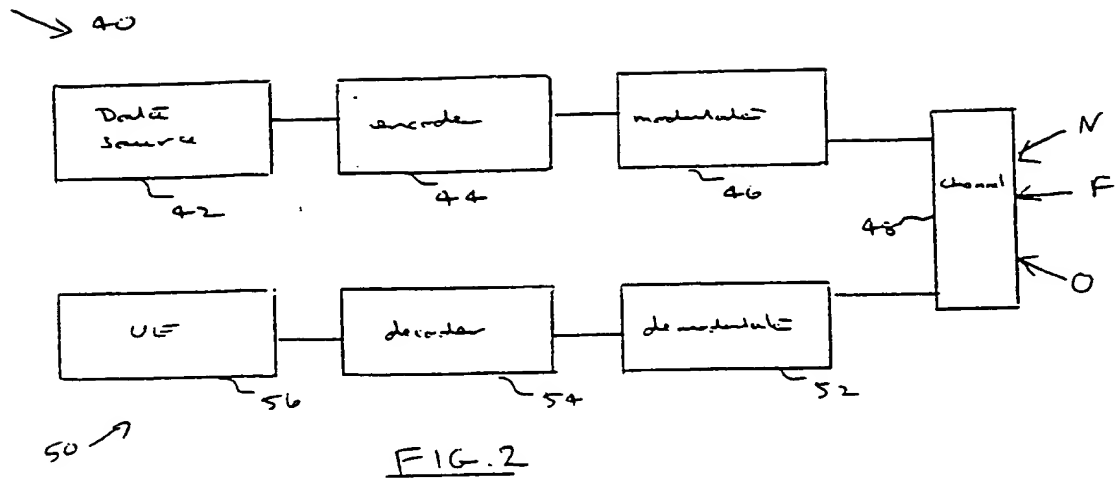
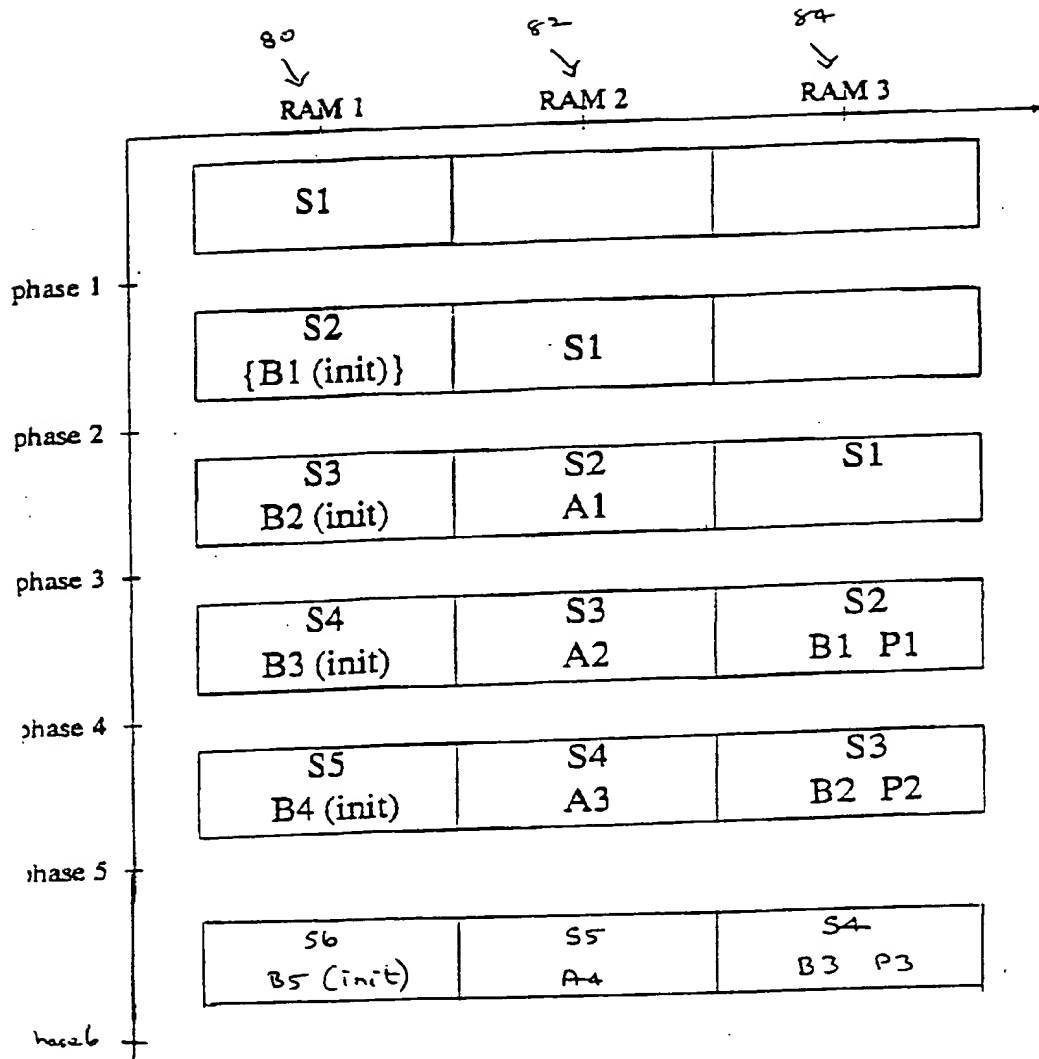


FIG. 1



FIG. 4

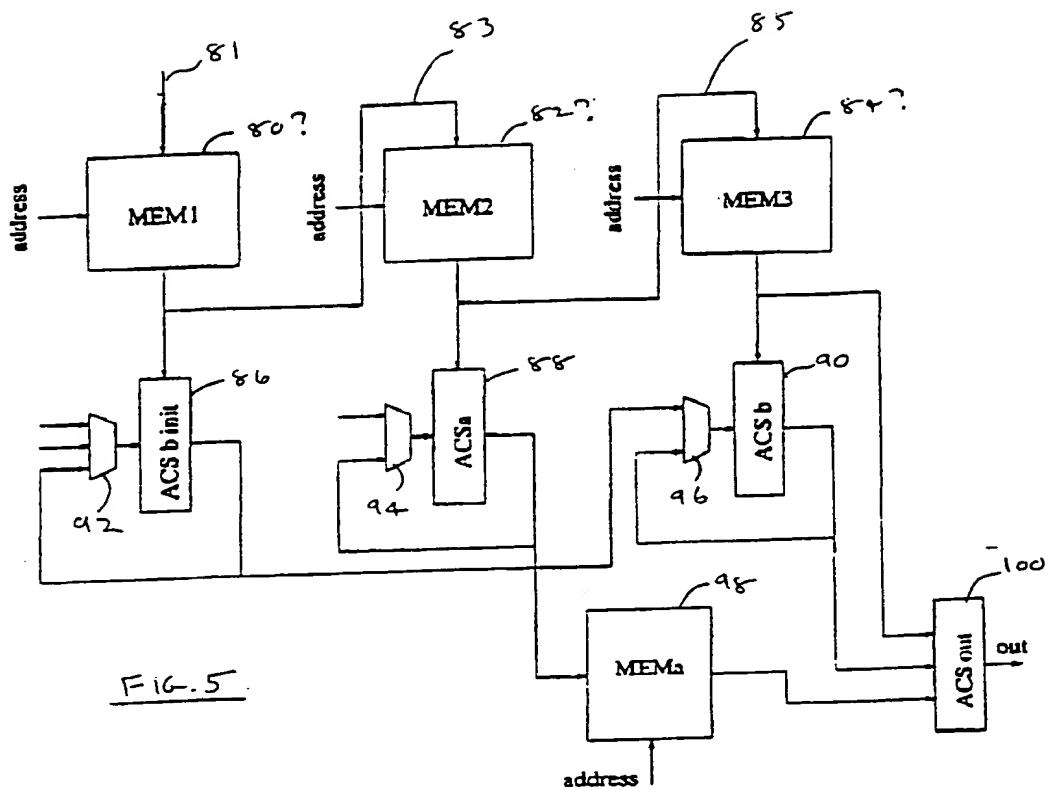


FIG. 5

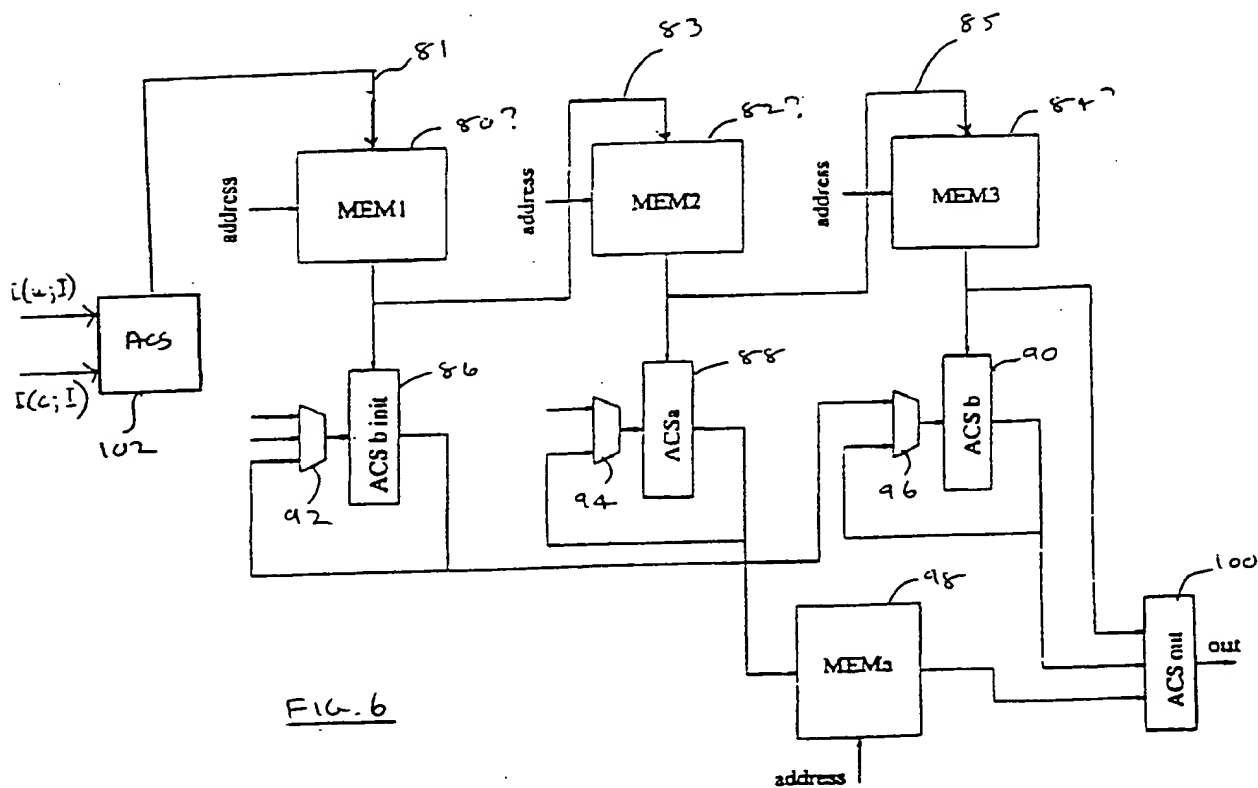


FIG. 6